JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2011.02362.x

Approximate Bayesian computation reveals the factors that influence genetic diversity and population structure of foxsnakes

J. R. ROW*, R. J. BROOKS†, C. A. MACKINNON†, A. LAWSON†, B. I. CROTHER‡, M. WHITE‡ & S. C. LOUGHEED*

*Department of Biology, Queen's University, Kingston, ON, Canada
†Department of Zoology, University of Guelph, Guelph, ON, Canada
‡Department of Biology, University of Biology, Southeastern Louisiana University, Hammond, LA, USA

Keywords:

assignment test;
cytochrome *b*;
glaciation;
microsatellites;
North America;
phylogeography;
population genetics;
reptile;
simulations.

Abstract

Contemporary geographical range and patterns of genetic diversity within species reflect complex interactions between multiple factors acting across spatial and temporal scales, and it is notoriously difficult to disentangle causation. Here, we quantify patterns of genetic diversity and genetic population structure using mitochondrial DNA sequences (101 individuals, cytochrome *b*) and microsatellites (816 individuals, 12 loci) and use Approximate Bayesian computation methods to test competing models of the demographic history of eastern and western foxsnakes. Our analyses indicate that post-glacial colonization and past population declines, probably caused by the infilling of deciduous forest and cooler temperatures since the mid-Holocene, largely underpin large-scale genetic patterns for foxsnakes. At finer geographical scales, our results point to more recent anthropogenic habitat loss as having accentuated genetic population structure by causing further declines and fragmentation.

Introduction

Identifying and quantifying the mechanisms underpinning geographical variation within species are fundamental to our most basic understanding of evolution and ultimately speciation (Gould & Johnston, 1972). Contemporary patterns of genetic diversity, and the distributions of organisms themselves, reflect both the legacy of historical factors like Pleistocene range fragmentation (e.g. Schoville & Roderick, 2009; Aldenhoven *et al.*, 2010; Qu *et al.*, 2010) and past demographic changes (e.g. population and range expansion and population bottlenecks; Austin *et al.*, 2002; Howes *et al.*, 2006) coupled with the influence of more recent factors (e.g. human-caused range fragmentation and isolation; e.g. Dyer *et al.*, 2010). Indeed, even distinguishing between what constitutes 'historical' and 'contemporary' is fraught with difficulty and distinction between the two terms is inconsistent in the literature (Eckert *et al.*, 2008).

Correspondence: Stephen C. Lougheed, Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.
Tel.: +1 613 533 6138; fax: +1 613 533 6617;
e-mail: steve.lougheed@queensu.ca

Regardless of the widespread recognition that all of these factors may play important roles in shaping contemporary genetic patterns, it has proved challenging to disentangle their respective contributions (Zellmer & Knowles, 2009). Because of these difficulties, traditional phylogeographical approaches that infer the relative contributions of past population processes through *post hoc* tests of an association between genealogical patterns and geography (Avise *et al.*, 1987; Templeton, 1998) can lead to spuriously attributing causation (Panchal & Beaumont, 2007).

There has been a recent move to embed phylogeography within a more rigorous hypothesis-testing framework, which allows for both the tests of competing models that are articulated *a priori* and formal tests of certainty (Knowles & Maddison, 2002; Beaumont *et al.*, 2010; Knowles & Alvarado-Serrano, 2010). Approximate Bayesian computation (ABC) coupled with coalescent modelling in population genetics (Beaumont *et al.*, 2002) is a promising method to accomplish this (Beaumont, 2010; Bertorelle *et al.*, 2010; Csillery *et al.*, 2010). As with all Bayesian analysis, prior information can be incorporated in the form of prior distributions, and competing models can be compared (Leuenberger & Wegmann,

2010). These characteristics combined with the ability to test complex and more realistic demographic scenarios (Bertorelle *et al.*, 2010) make it ideal for phylogeography. Although ABC approaches for population genetic and phylogeographical questions are underutilized (Beaumont *et al.*, 2002; Bertorelle *et al.*, 2010), they have already proven versatile and have been applied to test alternate demographic (Ray *et al.*, 2010) and evolutionary models (Fagundes *et al.*, 2007) and also to estimate demographic parameters (e.g. Estoup & Clegg, 2003; Wegmann & Excoffier, 2010).

Here, we use an ABC approach to test competing models of the demographic history of a North American temperate reptile, the foxsnake (*Pantherophis gloydi* and *P. vulpina*). The current northern range of foxsnakes is unusual among terrestrial squamates as it was almost completely covered by ice sheets approximately 100 000 years ago and was not free of ice until 10 000 years ago (Ehlers & Gibbard, 2004). A large, contemporary geographical range disjunction (see: Conant & Collins, 1991; Fig. 1) has produced some speculation over its cause (Morse, 1902; Schmidt, 1938) and has had implications for taxonomy (Conant, 1940; Collins, 1991). Populations on the eastern and western side of the disjunction are currently recognized as different species, the eastern foxsnake and the western foxsnake, respectively (Collins, 1991; Collins & Taggart, 2008). Western foxsnakes are prairie specialists (Conant & Collins, 1991), whereas eastern foxsnakes use both marsh and prairie habitat and occupy disjunct populations along the shorelines of the Great Lakes (Fig. 1; Row *et al.*, 2010). It has been suggested that, along with other species with similar habitat preferences, the eastern portion of their range resulted from an expansion following an eastward post-glacial steppe, which would provide open canopy habitat for prairie species (Schmidt, 1938; Smith, 1957). Pollen profiles suggest that the maximum extent of the post-glacial steppe conditions corresponded to the Hypsithermal period (~5000–7000 years ago) when temperatures were at a maximum during the Holocene (King, 1981; Webb, 1981). If the post-glacial steppe was responsible for the current eastern extension of their range, the return of deciduous forest combined with cooler temperatures could have subsequently caused local extinctions producing these large geographical disjunctions. It is also plausible that the extensive habitat fragmentation due to urban and agricultural development has caused or accentuated some gaps in distribution.

Here, we hypothesize that the fragmented regional eastern foxsnake populations represent relicts from the mid-Holocene when populations were larger and more connected due to the post-glacial steppe and the warmer temperatures of the Hypsithermal. Alternatively, one could also postulate that these populations were founded through dispersal events and also during the favourable conditions of the mid-Holocene. To test these competing hypotheses, we first quantify the patterns of genetic diversity and genetic population structure of foxsnakes using both mitochondrial and microsatellite DNA markers. We subsequently use ABC analysis to compare competing population demographic models that are consistent with these two different hypotheses: (i) large populations that have undergone drops in population size and splitting events and (ii) small founding populations that have split from large populations and expanded to be stable. We predict that due to the improbability of long distance dispersal events of snakes, models consistent with hypothesis 1 will have greater support. Because it is also possible for European settlement to be responsible for population declines and splitting events, we include this scenario in the priors of our demographic models for hypothesis 1 and determine the more likely model by examining the parameter values. Because of the complexity of some of the proposed models, we take a hierarchical approach throughout the ABC analysis,

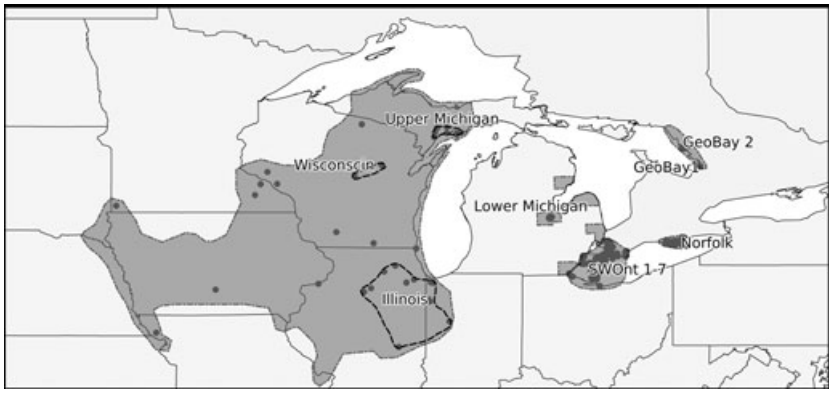

Fig. 1 Current approximate range of foxsnakes (dark grey) based on Ernst and Barbour (1989) and occurrence records from Michigan and Ontario. Grey dots represent the locations of one or more samples used in the analyses. Dashed lines circumscribe western foxsnake locations that we pooled for genetic diversity and differentiation analysis.

first focusing on single regional populations, then building to ultimately include models encompassing the entire foxsnake range.

Materials and methods

Genetic sampling

In 2006–2009, over the season when foxsnakes are active (May–September), we assembled 833 samples across the range of each taxon (70 western foxsnakes; 746 eastern foxsnakes). Samples collected by us or by researchers working in other regions were small blood samples (~200 mL stored in 95% ethanol) taken from the caudal vein of hand-captured individuals or from tissues samples (muscle and skin) collected from road kills. Eastern foxsnakes occur in four geographically disjunct populations (Fig. 1). For population-based analysis, samples were therefore organized into five regional populations: four corresponding to the eastern foxsnake disjunct populations [southwestern Ontario, lower Michigan (Shiawassee National Wildlife Refuge), Georgian Bay and Norfolk county, ON, Canada] and one corresponding to western foxsnakes where populations appear to be more continuously distributed. We further divided the Georgian Bay (Georgian Bay 1 and Georgian Bay 2) and western foxsnakes (Illinois; Wisconsin and upper Michigan) into local populations where we had geographical clusters of samples (Table S1; Fig. 1). Southwestern Ontario was divided into seven local populations (southwestern Ontario1 – southwestern Ontario7) based on the results of previous spatial assignment tests and strong differentiation across this region (Table S1; Row *et al.*, 2010).

We extracted DNA from blood and tissue using QIAGEN (Hilden, Germany) DNeasy blood and tissue kit following manufacturer's protocols. All samples were genotyped for 12 microsatellite loci (FS24, FS50, FS33, FS52, FS67, FS82, FS77, FS63, FS09B, FS42B, FSV16B and EOB10) (Blouin-Demers & Gibbs, 2003; Row *et al.*, 2008) following the methods outlined by Row *et al.* (2008) and Dileo *et al.* (2010). Deviations from neither Hardy–Weinberg equilibrium (HWE) nor linkage disequilibrium (Row *et al.*, 2008) were evident nor were null alleles prevalent (Dileo *et al.*, 2010; Row *et al.*, 2010). We also sequenced 101 individuals for 700 bp of mtDNA (cytochrome *b* region), but because of very low variation across most of our sampling range, we excluded these data from the ABC analysis but quantify genetic structure and diversity and provide the methodology and results in Supporting information (Appendix S1).

Microsatellite structure and diversity

We first quantified genetic population structure using assignment tests, which probabilistically assign individuals to their population of origin based on Hardy–Weinberg equilibrium and linkage equilibrium (reviewed in: Manel *et al.*, 2005). The number of genotyped samples for eastern foxsnakes outweighed our western foxsnake samples. To minimize the impact of this difference in sampling, we subsampled our eastern foxsnake samples and included only 10 random samples per local population when we had large sample numbers. This subsampling leads to a dataset comprised of 134 eastern foxsnake and 70 western foxsnake samples, which we used in a nonspatial admixture analysis in STRUCTURE 2.3.3 (Pritchard *et al.*, 2000). We ran 200 000 (100 000 burn-in) MCMC iterations 100 times for each of $k = 1$ to $k = 10$ using correlated allele frequencies and default parameters. The top 10 models for each k were averaged in CLUMPP 1.2 (Jakobsson & Rosenberg, 2007) and displayed using DISTRUCT 1.1 (Rosenberg, 2004). We examined the variation in log-likelihood for the top 10 runs for each value of k to ensure that run length and burn-in were sufficient.

We determined differentiation between defined populations (Fig. 1) using pairwise F_{ST} (Weir & Cockerham, 1984) and Jost's D differentiation (Jost, 2008). Within defined populations, we calculated expected heterozygosity (H_e – corrected for sample size; Nei, 1978), mean number of alleles, mean F_{IS} and standardized allelic richness (Hurlbert, 1971) using microsatellite analyser 4.05 (Dieringer & Schlotterer, 2003). We determined whether there were significant differences between populations using a Friedman test, which is a nonparametric repeated measures analysis, where our measure of diversity was ranked for each group (population) within the blocks (loci), and significance was determined with a χ^2 test (df = number of groups − 1). We did pairwise Wilcoxon–Nemenyi–McDonald–Thompson *post hoc* tests of differences between populations (Zar, 1996; Hollander & Wolfe, 1999) with sequential Bonferroni correction (Rice, 1989).

Demographic modelling with approximate Bayesian computation

For ABC analysis, genetic datasets are generated from coalescent simulations using population parameters, drawn from a prior distribution, under a specified model. For each simulation, summary statistics (e.g. allelic range, number of alleles and F_{ST}) are calculated, and the Euclidean distance (using the multivariate space of the summary statistics) between the generated and actual summary statistics is calculated. Models can be compared and parameters estimated by retaining a proportion of the simulations with the lowest Euclidean distance to the observed summary statistics (e.g. Ray *et al.*, 2010) or the simulations that are below (in Euclidean distance) a set threshold (e.g. Fagundes *et al.*, 2007). We used ABCtoolbox (Wegmann *et al.*, 2010), which consists of four programs: SIMCOAL 2.0 (Laval & Excoffier, 2004), arlsumstat (Excoffier & Lischer, 2010), ABCsampler and

ABCestimator. Together these programs: (i) generate coalescent simulations, (ii) calculate summary statistics, (iii) calculate standardized (common mean and standard deviation) Euclidean distances and retain the generated simulations closest to the actual dataset and (iv) perform a post-sampling regression adjustment to estimate the posterior distribution for parameter estimation. When included as a summary statistics, we used a modified python script of SMOGD version 1.2.5 (Crawford, 2010) to calculate pairwise Jost's D.

For the simulations, we set the number of loci and sample sizes to those of the actual dataset and microsatellite diversity was generated under a strict stepwise mutational model (SMM). Because two microsatellite loci (EOB10 and FS09) had large gaps in repeat number implying marked departure from a strict SMM, they were excluded from the ABC analysis and a maximum of 50 individuals were chosen from any given population to reduce computing time. Unless stated otherwise, 5×10^5 simulations were run for each model, and the 5000 simulations with the lowest Euclidean distance were retained for model testing and parameter estimation.

Because large-scale demographic models that include all populations have large numbers of parameters making calculations computationally intensive, we used a hierarchical approach by first modelling regional populations separately to allow us to more confidently fix or narrow the range of priors for parameters in the range-wide models. It is possible, however, for population structure to mimic bottlenecks (Nielsen & Beaumont, 2009; Chikhi *et al.*, 2010), which could potentially lead to spurious results when regional population models are considered alone. We therefore compared the selected models and estimated parameters from the regional and range-wide models to ensure consistency when considering models at different scales.

Model descriptions and model parameters are described in the following sections. For the Georgian Bay region, we had samples from two locations separated by ~50 km. To simplify the models, for all ABC analyses, we only used samples from the more southerly population (Georgian Bay 1; Fig. 1), where we had a greater sample size.

Model choice

Following the selection of the datasets with the lowest Euclidean distance to the observed summary statistics, we estimated the fit and compared competing historical-demographic models using three different methods. First, we used ABCtoolbox to calculate the marginal density for each of the retained 5000 simulations and calculate a *P* value as the fraction of these that have a lower marginal density. A low *P* value indicates that most retained simulations have higher marginal densities and suggests an inability of the model to produce the observed summary statistics (Wegmann *et al.*, 2010). Second, we calculated the Bayes factor as the probability of one model versus another (Wegmann *et al.*, 2009, 2010). Third, following Pritchard *et al.* (1999), for each model, we combined the 5000 simulations with the lowest standardized Euclidean distances in multivariate space (15 000 total), and then estimated the relative probability of each model as the proportion of simulations that were included in the top 1000 models (of the 15 000) with the lowest overall Euclidean distances.

Parameter estimation and model checking

To estimate population parameters, we applied a general linear model (ABC-GLM) post-sampling regression adjustment to the 5000 retained simulations (Leuenberger & Wegmann, 2010), as implemented in ABCestimator. The regression adjustment assumes a linear model within a narrowed prior based on the retained simulations, and calculates the density of 100 evenly spaced points along the parameter value, to generate the posterior distribution. We report the mode and 90% highest posterior density (HPD) interval as an estimate of that population parameter. To convert splitting times from generations to years, we assumed a generation time of 7.5 years (half way between age at maturity and longevity averaged for Georgian Bay and southwestern Ontario; J.R. Row and S.C. Lougheed, unpublished).

The potential of the parameter to be correctly estimated by the summary statistics was summarized by calculating the coefficient of determination (R^2) for a multiple regression of the parameter against all summary statistics, using all of the simulated datasets (Neuenschwander *et al.*, 2008; Ray *et al.*, 2010). Neuenschwander *et al.* (2008) suggested that parameters with an R^2 of < 10% are unreliable, because the summary statistics explain little of their variability.

For the selected models, we also estimated fit and model checked by examining the posterior predictive distribution (Cornuet *et al.*, 2010; Csillery *et al.*, 2010). We generated 10 000 simulations under the selected model with the priors set to the 90% HPD estimates. Subsequently, we ran principal components analysis (PCA) on the resulting summary statistics and determined whether the observed data fell within the posterior predictive distribution by visually inspecting plots of the first 10-PCA component (Cornuet *et al.*, 2010).

Population-scale analysis

For the three single eastern foxsnake regional populations (lower Michigan, Georgian Bay and Norfolk County) and the Illinois population, we compared three different demographic models: (i) *Drop* – large populations that underwent instantaneous drops to their current size, (ii) *Decline* – large populations that underwent exponential declines to their current size and (iii) *Bottle* – small founding populations that expanded to their present-day stable configurations (Fig. 2a). We refined prior distributions by testing models with different prior distributions and comparing the marginal density

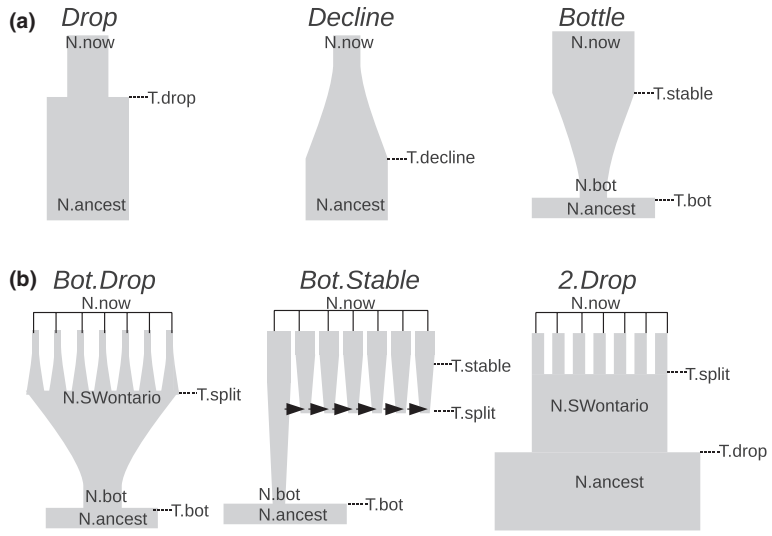

Fig. 2 Population demographic models used in Approximate Bayesian computation analysis for (a) single populations (Illinois, Georgian Bay 1, lower Michigan and Norfolk; Fig. 1) and (b) southwestern Ontario, where a number of genetic clusters have been identified (Row *et al.*, 2010). Additional details of models and parameters [T.Drop, time (generations) of population drop; T.decline, time (generations) of exponential decline; T.stable, time (generations) since population has become stable; T.split, time (generations) of population split; N.Now, current effective population size (N_e); N.SWOntario, N_e of combined population in southwestern Ontario; N.Ancest, ancestral N_e; N.bot, N_e of population bottleneck] are in the text.

between models (Table S2). In the *Drop* and *Decline* population models, prior distributions on the timing of the drop or decline were wide enough to accommodate the possibility that reduction in population size resulted either from reforestation and from cooler temperatures after the Hypsithermal (~2000–8000 years before present) or from human habitat loss and fragmentation (10–150 years before present). In the expansion models, priors on the timing of the founding event included the Hypsithermal, when increased temperatures and the post-glacial steppe conditions would have been optimal for foxsnakes. Separate priors were used for eastern and western foxsnake populations because of different expectations (e.g. expect southwestern populations to have been established further in the past as predicted by a south to north post-glacial colonization history) and marginal densities during testing. For this smaller scale analysis, we used the mean and standard deviation (calculated over loci) for four summary statistics: number of alleles, expected heterozygosity, modified Garza–Williamson index (Garza & Williamson, 2001; Excoffier *et al.*, 2005) and allelic size range, thus a total of eight statistics for model comparison.

Because southwestern Ontario is comprised of seven genetic clusters identified previously through spatial assignment tests (Row *et al.*, 2010), for it we tested more complex models representing alternative possible demographic histories (Fig. 2b, Table S3): (i) *Bot.Drop* – a small population founded southwestern Ontario, expanded to a large population, then subsequently split into seven populations, which all underwent exponential decline into their current population sizes, (ii) *Bot.Stable* – the seven populations were sequentially colonized from a small population and then exponentially expanded into stable populations and (iii) *2.Drop* – a large population dropped to a smaller population, (i.e. split from the other regional populations) and then dropped and split into the current seven populations. For each population, we used the eight summary statistics listed above, but added pairwise Jost's D (Jost, 2008) as a metric of differentiation between populations. With the inclusion of seven populations and Jost's D, the total number of summary statistics is large (77 statistics), which can lead to statistical noise and make posterior parameter estimation difficult (Joyce & Marjoram, 2008). Following the methods of Wegmann *et al.* (2009) and using an R (R Development Core Team, 2009) script provided with ABCtoolbox, we therefore reduced the statistical summary space using a partial least squares (PLS) approach to include uncorrelated orthogonal components that explain the largest amount of variation in the parameter set. The number of PLS components to include was chosen by visually determining when additional components did not reduce the root mean square error of the parameters.

Range-wide scale

At the regional scale, the Wisconsin and upper Michigan populations were combined because they had low sample sizes, and the STRUCTURE analysis suggested that they belonged to the same genetic cluster. We tested three models that included 12 populations (Illinois, Wisconsin

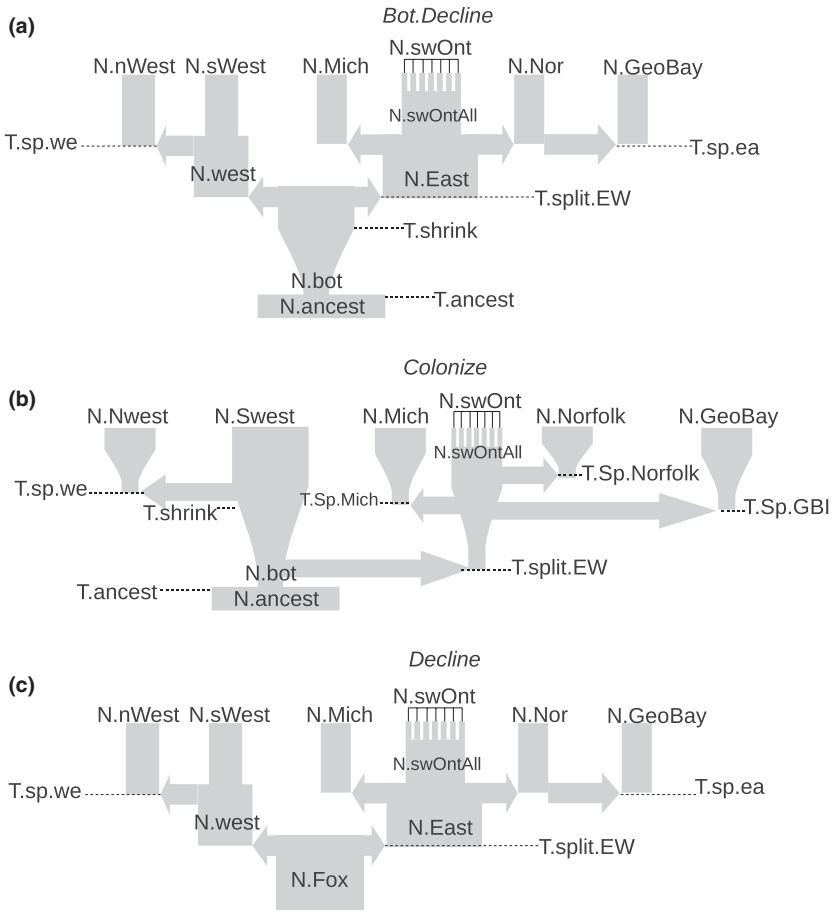

Fig. 3 Three possible colonization models of foxsnakes into their current range and used in the approximate Bayesian computation analysis. Additional details of models and parameters are in the text.

& upper Michigan, southwestern Ontario 1–7, lower Michigan, Norfolk and Georgian Bay 1). To reduce complexity, southwestern Ontario population sizes were set to be gamma distributed as Gamma (8,8/X), where 'X' was the average of southwestern Ontario population sizes. Using the gamma distribution with the prior distribution of the mean (400–2000), the population sizes and population size variation observed in the small-scale southwestern Ontario population models were possible. The merging of the southwestern Ontario regional population was set to 10–80 generations, which was not set to match the values found in the southwestern Ontario population model, but rather allowed recent coalescent events to occur within each population (Ray *et al.*, 2010).

At this large scale, we tested three different models that we think best reflect possible historical demographic scenarios based on our current knowledge of regional post-Pleistocene and the species' ecology: (i) *Bot.Decline* – after a population bottleneck (i.e. in a glacial refugium) foxsnakes expanded exponentially to a large population representing their current range. Consistent with the forest infill hypothesis, populations then began to drop and fragment (Fig. 3a; Table S4). (ii) *Colonize* – after a population bottleneck, foxsnakes populations colonized their current range through sequential founder populations and subsequent population expansions (Fig. 3b; Table S4). (iii) *Decline* – a variant of the *Bot.Decline* model where there is no initial bottleneck for foxsnake populations (Fig. 3c; Table S4). Summary statistics were the same as for the southwestern Ontario population models. Because we were not attempting to estimate divergence times of the southwestern Ontario populations in this model, we combined the local southwestern Ontario populations into one regional population before calculating the pairwise Jost's D.

Last, because the range-wide models were complex and we may not have the power (particularly for western foxsnakes) to discern finer scale patterns, we tested a simple two-population model by separately pooling the eastern and western populations together and recalculating the observed summary statistics (same summary statistics as for the complex models, but for only two populations). Under this simple model, the two

populations (N_e: 2000–200 000) diverged at some point in the past (10–2000 generations), from a large ancestral population (N_e: 20 000–400 000). We used the same microsatellite mutation parameters as in the complex models. Because of the different number of summary statistics, we could not directly compare this model with the complex models but compared the estimated parameters (e.g. splitting times and population sizes).

Effect of gene flow

Because of the complexity of our models and the large number of parameters, we did not include gene flow. This is obviously unrealistic, and to address the impacts of this necessary simplification in our models, we used simulations to test the effect of past gene flow on parameter estimation (Appendix S3).

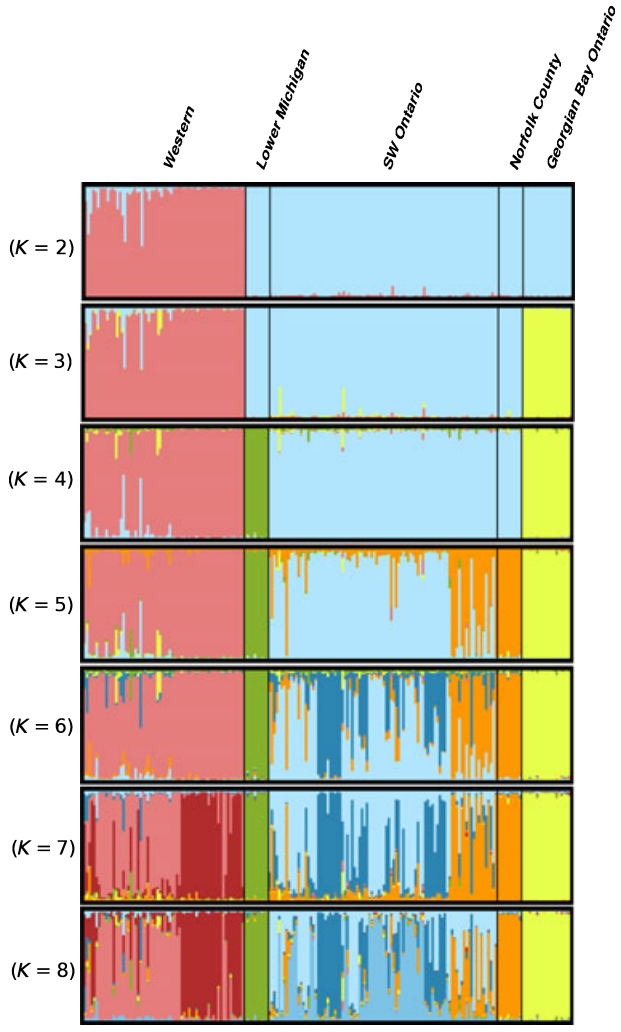

Fig. 4 Bar plots representing admixture coefficients for eastern and western foxsnakes from assignment test analyses performed in STRUCTURE 2.3.3. The top 10 runs (highest log probability of data) from 100 replicates were averaged in CLUMPP 1.2 and displayed with DISTRUCT 1.1 for each of $k = 2$ through $k = 8$. See Fig. 1 and text for description of populations.

Results

Microsatellite structure and diversity

For our Bayesian assignment tests, the log probability of the data reached a plateau around $k = 6$ to $k = 8$, suggesting that the most likely number of clusters was within that range. The top 10 of 100 runs (based on highest log probability of data) for each of $k = 2$ through $k = 8$ were averaged in CLUMPP 1.2 (Jakobsson & Rosenberg, 2007) and displayed with DISTRUCT 1.1 (Fig. 4). The first major split ($k = 2$) was between eastern and western foxsnakes. This division remained for all values of k and suggested very little admixture between western foxsnakes and any of the eastern foxsnake populations. Overall, there was clearly more genetic structure within eastern foxsnakes, with both the Georgian Bay (at $k = 3$) and lower Michigan regional populations (at $k = 4$) separating from the other populations with little suggested admixture. The remaining clusters (defined at $k = 5$ through 8) were less clear with some admixture between the Norfolk regional population and the other southwestern Ontario populations. The appearance of additional clusters within southwestern Ontario and within western foxsnakes was present when k was set to the highest values ($k = 7$ & $k = 8$). The divisions revealed by STRUCTURE were also supported by a PCA on the microsatellite genotypes conducted with the *adegenet* package (Jombart, 2008) in R (R Development Core Team, 2009) (Fig. S1).

Pairwise F_{ST} values ranged from 0.07 to 0.10, between western foxsnake populations and from 0.05 to 0.60 between eastern foxsnake populations (Table S5). Between eastern and western populations pairwise F_{ST} values were between 0.10 and 0.61 (Table S5), with the Illinois population most similar to eastern foxsnakes. The patterns with Jost's D differentiation mirrored those of F_{ST}. All pairwise F_{ST} values were significantly different from zero ($P < 0.001$).

F_{IS} values were not significantly different among defined local populations ($\chi^2_{13} = 18.93$, $P = 0.12$), but both allelic richness ($\chi^2_{13} = 73.64$, $P < 0.001$) and H_e ($\chi^2_{13} = 42.52$, $P < 0.001$) values varied significantly among populations. The three most isolated eastern foxsnake populations (Georgian Bay 1&2; lower Michigan; and Norfolk) had the lowest allelic richness estimates, and these were significantly lower than the Illinois population when compared using a Wilcoxon–Nemenyi–McDonald–Thompson *post hoc* test (Zar, 1996; Hollander & Wolfe, 1999) (Table 1). Similarly, H_e was lowest in the three isolated regional populations, but only two (Georgian Bay 1 and Norfolk) were significantly different from the population with the highest H_e (Illinois).

Table 1 Sample size (N), mean expected heterozygosity (H_e), mean number of alleles (MNA) and mean allelic richness (AR) for local populations of foxsnakes (Fig. 1). Standard deviation is given in brackets and populations connoted with different superscript letters for H_e and allelic richness were significantly different. F_{IS} was not significantly different between populations and MNA was not tested. See text for details of tests and Fig. 1 for distribution of populations.

Population	N	H_e	MNA	AR	F_{IS}
GeoBay1	119	0.28 (0.13)^{b}	2.41 (0.79)	2.05 (0.54)^{b}	0.03 (0.09)
GeoBay2	41	0.36 (0.21)^{ab}	2.81 (0.75)	2.40 (0.66)^{b}	−0.02 (0.13)
Swont1	62	0.59 (0.14)^{ab}	4.33 (1.61)	3.69 (1.07)^{ab}	0.01 (0.13)
SWont2	134	0.61 (0.13)^{ab}	5.50 (1.83)	3.97 (1.10)^{ab}	0.04 (0.05)
SWont3	28	0.52 (0.14)^{ab}	4.00 (1.13)	3.46 (0.75)^{ab}	0.05 (0.16)
SWont4	142	0.53 (0.20)^{ab}	4.91 (1.73)	3.69 (1.27)^{ab}	0.02 (0.05)
SWont5	28	0.58 (0.15)^{ab}	3.83 (1.40)	3.43 (1.06)^{ab}	−0.01 (0.16)
SWont6	84	0.62 (0.11)^{ab}	5.33 (1.72)	3.93 (0.87)^{ab}	0.12 (0.06)
SWont7	47	0.50 (0.16)^{ab}	4.08 (1.50)	3.23 (0.94)^{ab}	0.03 (0.17)
Norfolk	64	0.32 (0.19)^{b}	3.25 (1.28)	2.51 (0.80)^{b}	0.13 (0.11)
L. Mich	33	0.45 (0.22)^{ab}	2.08 (1.08)	2.04 (1.04)^{b}	0.02 (0.18)
Illinois	27	0.74 (0.12)^{ab}	7.25 (1.76)	5.96 (1.44)^{a}	0.07 (0.11)
Wisc.	12	0.61 (0.19)^{ab}	4.33 (1.40)	4.33 (1.61)^{a}	0.03 (0.14)
U. Mich	12	0.55 (0.25)^{ab}	3.83 (1.80)	3.80 (1.75)^{a}	0.01 (0.18)

Table 2 Comparison of Approximate Bayesian computation models using marginal densities, probabilities (low P value indicates an inability of the model to produce the summary statistics) and relative probabilities. Models presented in Figs 2 and 3 and described in more detail in the text.

Population	# PLS*	Model	Mar. Density	P value	Rel. Prob.
Michigan	NA	*Drop*	2.6×10^{-1}	0.71	0.74
Michigan	NA	*Decline*	9.9×10^{-2}	0.71	0.26
Michigan	NA	*Bottle*	1.6×10^{-3}	0.41	0.02
GBI	NA	*Drop*	4.1×10^{-3}	0.69	0.59
GBI	NA	*Decline*	1.5×10^{-3}	0.15	0.39
GBI	NA	*Bottle*	1.3×10^{-5}	0.07	0.02
Norfolk	NA	*Drop*	1.7×10^{-4}	0.01	0.50
Norfolk	NA	*Decline*	7.9×10^{-6}	<0.001	0.46
Norfolk	NA	*Bottle*	1.7×10^{-6}	0.01	0.04
Illinois	NA	*Drop*	12.1	0.99	0.75
Illinois	NA	*Decline*	3.3	0.98	0.25
Illinois	NA	*Bottle*	6.7×10^{-5}	0.98	0
swOnt	10	*Bot.Decline*	5.1×10^{-5}	0.70	0.42
swOnt	10	*Bot.Stable*	1.3×10^{-5}	0.95	0.14
swOnt	10	*2.Drop*	2.4×10^{-4}	0.99	0.44
Full	15	*Bot.Decline*	1.3×10^{-12}	0.003	0.22
Full	15	*Colonize*	1.9×10^{-16}	<0.001	0
Full	15	*Decline*	2.9×10^{-12}	0.01	0.77

*Number of Partial Least Squares components, see text for details.

Table 3 Prior distribution and posterior probabilities (with 90% highest probability density (HPD)) for parameters of the *Drop* single population models (Fig. 2a). N parameters are effective population sizes (N_e) and T parameters are time in generations.

Parameter	Population	Mode	90% HPD	R^2
N.now	Michigan	774	100–1564	0.49
N.ancest	Michigan	140 606	57 454–200 000	0.37
T.drop	Michigan	430	100–860	0.31
N.now	GeoBay	392	100–978	0.49
N.ancest	GeoBay	166 343	75 273–200 000	0.38
T.drop	GeoBay	300	40–700	0.30
N.now	S.West	10 754	4722–17 988	0.57
N.ancest	S.West	94 948	46 464–175 758	0.29
T.drop	S.West	2093	734–2970	0.19

N.now, N_e; N.ancest, ancestral N_e; T.drop, time of decline in population size.

Demographic modelling with approximate Bayesian computation

Population-scale

For the lower Michigan, Georgian Bay and Illinois populations the marginal densities of all three models (*Drop*, *Decline* and *Bottle*) had P values above 0.05 (Table 2). This indicated that the observed marginal densities were within the range of the distribution of marginal densities for the retained simulations and capable of producing the observed summary statistics. The marginal densities of the *Drop* model, however, were highest for all three populations with Bayes factors of 3 and 163 for lower Michigan, 3 and 316 for Georgian Bay and 4 and 182 282 for Illinois, comparing the *Drop* model to the *Decline* and *Bottle* models, respectively. The marginal density for the Norfolk population had P values that were <0.05 for all of the models, suggesting that none of these models could accurately produce the summary statistics. Examining the posterior distributions for the *Drop* model for lower Michigan and Georgian Bay, it appears that both had a significant drop in effective population size (N_e) around 3225 and 2250 years in the past, respectively (Table 3). Current N_e was larger for lower Michigan (mode of 774 individuals) than for Georgian Bay (mode of 392 individuals) (Table 3). The decrease in N_e in Illinois appeared to occur much earlier (15 697 years in the past) and resulted in a larger current

N_e (10 754) (Table 3). The Illinois population stretches over a much larger area (Fig. 1), so these populations cannot be directly compared. The observed statistics were well within the posterior predictive distribution for all the significant single-population *Drop* models (Georgian Bay, lower Michigan, Illinois) (Appendix S2).

For southwestern Ontario, the marginal densities of all three models again had *P* values above 0.05 (Table 2). When comparing the marginal densities of the models, the *2.Drop* model had the highest *P* value resulting in Bayes factors of 18 and 5 compared with the *Bot.stable* and *Bot.decline* models, respectively. For the *2.Drop* model, the N_e posterior distributions had modes that ranged from 1236 to 3200 individuals and posterior distributions for the timing suggested a large drop in population size 15795 years (90% HPD: 9322–18 525) in the past and a drop/split of the southwestern Ontario populations 975 years (90% HPD: 150–1875) in the past (Table 4). The first N_e drop (T.Drop), however, had an R^2 value of much <10% suggesting our model was not able to predict this parameter with any accuracy. Plots of the PCA components of the posterior predictive distribution suggested this model could adequately produce the observed summary statistics (Appendix S2).

Range-wide scale

At the range-wide scale, the *Bot.Decline* and *Decline* models both had much higher marginal densities than the *Colonize* model, but there was conflicting evidence over which of these two models had stronger support (Table 2). The marginal density for the *Decline* model was higher leading to a modest Bayes factor of 2.23 when comparing the *Decline* to *Bot.Decline* model, but the relative probability (i.e. proportion of simulations within top 1000 simulations) was higher (0.77) for the *Bot.Decline* model. Neither of these models, however, had *P* values above 0.05, suggesting they could not reliably produce the observed summary statistics. This equivocal result could be due to the complexity of the models that we were testing and/or the low sample sizes and reduced sampling coverage for western foxsnakes. We therefore used the same generated dataset but reduced the complexity by only including eastern foxsnake parameters and summary statistics when calculating Euclidean distances between the generated and actual datasets, and in the post-sampling regression adjustment. After reducing the models, the *P* value for the *Decline* (*P* value = 0.09), but not the *Bot.Decline* (*P* value = 0.004) model was above 0.05. The marginal density was also higher for the *Decline* model resulting in a Bayes factor of 56.77, and we therefore only estimated the parameters of this simplified *Decline* model (Table 5).

The simplified *Decline* model suggested eastern foxsnake regional populations split approximately 2340 years ago (90% HPD: 750–4455), which matched well with the timing of the population drop for the lower Michigan and Georgian Bay single-population models, and would have been before major European settlement. The split between eastern and western foxsnakes was estimated at 9817 years (90% HPD: 5317–14 587). This is again well before European settlement, but also seems to predate the timing for the infilling of deciduous forest into this region. Relatively consistent with the single-population models, the N_e for the lower Michigan and Georgian Bay populations were 788 and 642 individuals, respectively. For southwestern Ontario, the mean population size (mode = 772) was lower than any of the

Table 4 Prior distribution and posterior probabilities (with 90% highest probability density (HPD) estimate) for parameters of the *2.Drop* model for southwestern Ontario (Fig. 2b). N parameters are effective population sizes (N_e) and T parameters are time in generations.

Parameter	Priors	Mode	90% HPD	R^2
N.SWont1	400–4000	1200	400–2254	0.69
N.SWont2	400–4000	3272	2036–4000	0.68
N.SWont3	400–4000	2908	1636–3962	0.69
N.SWont4	400–4000	1600	580–2836	0.68
N.SWont5	400–4000	3090	1818–3962	0.69
N.SWont6	400–4000	2000	762–3344	0.69
N.SWont7	400–4000	1636	618–3346	0.69
T.split	10–1000	130	20–250	0.73
T.Drop	1000–2500	2106	1243–2470	0.02
N.SWont	4000–40 000	6546	4000–14182	0.24
N.ancest	40 000–200 000	148 282	67 474–193 536	0.33

N.SWont1 to 7, N_e of southwestern Ontario populations of populations; N.ancest, N_e of ancestral population size before first drop; N.SWont, N_e of southwestern Ontario population before splitting; T.drop, time (generations) of first drop in population size; T.Split, time (generations) of second population drop and split into current populations.

Table 5 Prior distribution and posterior probabilities (with 90% highest probability density (HPD) estimate) for parameters of the simplified *Decline* regional model (see Fig. 3c and text for details). N parameters are effective population sizes (N_e) and T parameters are time in generations.

Parameter	Prior	Mode	90% HPD	R^2
N.swOnt	400–2000	772	400–1384	0.33
N.Mich	400–2000	788	416–1256	0.58
N.Norfolk	400–2000	1450	868–1918	0.58
N.GeoBay	400–2000	642	400–1046	0.58
N.swOntAll	10 000–100 000	19 986	10 000–55 398	0.15
N.East	10 000–100 000	33 756	10 162–78 220	0.10
N.Fox	20 000–200 000	158 182	50 910–200 000	0.40
T.split.EW	200–2000	1309	709–1945	0.16
T.sp.ea	50–1500	312	100–594	0.67

N.swOnt, N_e of southwestern Ontario populations; N.Mich, N_e of lower Michigan population; N.Norfolk, N_e of Norfolk population; N.GeoBay, N_e of Georgian Bay population; N.East, N_e of eastern foxsnake population before fragmenting; N.Fox, N_e of foxsnake populaton before splitting from western foxsnakes

population sizes estimated when we ran the southwestern Ontario alone. The gamma distribution [Gamma (8,8/X)], with an N_e mean of 772, allowed population sizes to vary between 200 and 1600; thus, when incorporating the 90% HPD (400–1384), the mean N_e would be well within the confidence intervals of the southwestern Ontario population model.

Because some of the parameters were not estimated in the simplified model (N_now_Swest; N_now_Nwest; N_West, T_Mer_West), the priors were left at their original values for these parameters, and their associated summary statistics were not included when calculating the posterior predictive distribution. Plots of the first 10-PCA components again suggested that the observed data were within the posterior predictive distribution (Appendix S2).

When combining eastern and western foxsnakes separately for the two-populations model, the N_e of eastern foxsnakes (Mode: 2000; 90% HPD: 2000–20 000) was estimated to be much lower than western foxsnakes (Mode: 24 000; 90% HPD: 2000–64 000). The mode of the eastern foxsnake N_e, however, was the same as the lower prior set for this parameter. The splitting time between eastern and western populations (Mode: 6555 years; 90% HPD: 1732–13 185) was less than with the complex models, but well within the confidence intervals.

Effect of gene flow

Simulations show that when gene flow occurred among simulated populations, but was not included in the ABC analysis, it generally neither led to model rejection nor had large effect on parameter estimation. Importantly, the largest effect was an over estimation of splitting times (i.e. if there is a splitting event with gene flow in actuality, but gene flow is not incorporated into the ABC analyses, then the estimate we derive from simulations ignoring gene flow is closer to the present; Appendix S3).

Discussion

Our microsatellite analysis revealed that genetic population structure and differentiation were greater in eastern foxsnakes, and genetic diversity was lower in isolated peripheral populations. Model comparisons suggested that these patterns were the result of large drops in population size, combined with past population divisions. Given the estimated timing of population size drops and population splits at the largest spatial scale, the most likely cause was the infilling of deciduous forest and/or cooler temperatures since the Hypisthermal and not habitat changes resulting from human alterations on the landscape. Further population declines and fragmentation in southwestern Ontario, however, were also evident and most likely caused by anthropogenic habitat loss and fragmentation.

Sequence data from the cytochrome *b* region of mtDNA showed minimal variation and patterns were not consistent with nuclear DNA analysis nor corresponded with the current distribution of fragmented populations. All but two eastern foxsnakes (58 total) and most of the western foxsnakes, east of the Mississippi, had an identical haplotype.

Genetic diversity and Genetic population structure

Assignment tests identified a clear split between the currently recognized ranges of eastern and western foxsnakes, with genetic structure more pronounced within eastern foxsnakes. This was expected given that the distribution of eastern foxsnakes appears to be more fragmented but suggests a more continuous distribution for western foxsnakes likely relating to more continuous habitat distribution or differences in habitat preference between the species. This fragmentation and geographical isolation has impacted microsatellite diversity; the isolated eastern foxsnake regional populations had significantly lower expected heterozygosity and allelic richness than the Illinois foxsnake population. Other studies on temperate species have found that, as populations move northwards away from glacial refugia, there is a decrease in genetic diversity (Johansson *et al.*, 2006; Howes & Lougheed, 2008). For foxsnakes, this difference in genetic diversity was nonsignificant between the Illinois population and the Wisconsin and upper Michigan populations. The upper Michigan population is likely at least as far from potential glacial refugia as some of the isolated eastern foxsnake populations. Because western foxsnake populations appear to be more continuously distributed, this lack of decline may be attributed to ongoing gene flow with southern populations, which would contribute to the maintenance of genetic diversity (Wright, 1978; Slatkin, 1987).

We found two major mtDNA clades (1.5% divergence), but in contrast with the microsatellite analysis, they did not correspond to the current designation of eastern and western foxsnakes or to any of the eastern foxsnake regional populations (Appendix S1). The Mississippi river was a barrier to gene flow for other species (Burbrink *et al.*, 2000; Howes *et al.*, 2006), and based on the distribution of haplotypes (Appendix S1), it is possible that this was a historical barrier for foxsnakes that has recently been transgressed. In the Wisconsin population (Fig. 1), we found haplotypes from both mt. clades, but microsatellite assignment tests put these individuals in the same genetic cluster, implying that the lineages are not reproductively isolated. Other snake species with similar divergences between cytochrome *b* lineages also show no evidence of assortative mating in zones of contact indirectly implying lack of reproductive isolation between clades (Gibbs *et al.*, 2006). Overall, the mtDNA diversity was low, and the majority of foxsnakes, east of the Mississippi (including all but 2 eastern

foxsnakes), had a single haloptype suggesting a bottleneck or selective sweep prior to the split between eastern and western foxsnakes. The cytochrome *b* region of mtDNA has been found to be variable and informative for closely related snake species (Burbrink *et al.*, 2000) suggesting that this paucity of diversity in eastern foxsnakes was not an artefact of the region of mtDNA examined.

Colonization patterns and approximate Bayesian computation analysis

We are aware that, although our models were relatively complex, there remained many simplifications (e.g. no gene flow and combined eastern foxsnake splitting times). The exclusion of gene flow in our models had little impact on population size estimates (Appendix S2) but could result in shifting the estimates of divergence time if strong gene flow was present. Overall models (both single–population models and regional models) that included large population declines consistently had better support than models with founder effects and subsequent population expansion. The current geographical range of foxsnakes was mostly covered by ice sheets approximately 100 000 years ago, and so there is no doubt that ancestral populations expanded into their current range since that time. Our models suggested, however, that ancestral foxsnake populations were once larger and more widely distributed across this region and that subsequent declines and population fragmentation have had the largest effect in shaping the current genetic diversity and structure. Based on herpetofauna distribution patterns, Schmidt (1938) suggested that a post-glacial steppe extended prairie like conditions eastwards from the prairie peninsula, which has been backed up by pollen profiles (King, 1981; Webb, 1981). These prairie conditions combined with the higher temperatures during the climatic optimum ~5000 years ago (Smith, 1957; Churcher & Karrow, 2008) arguably permitted these higher population sizes and/or greater connectivity across the range of eastern and western foxsnakes.

The maximum eastwards extent of prairie conditions has been estimated at approximately 5000–7000 years ago with subsequent westwards retreat until approximately 2000 years ago (Webb, 1981). Based on our complex models, the fragmentation of eastern foxsnakes populations occurred approximately 2340 years in the past (90% HPD confidence interval of 750–4455). This estimate precludes the possibility that the major geographical disjunctions within eastern foxsnakes were caused by European settlement and would be consistent with the infilling of deciduous forest and post-Hypsithermal cooling of temperatures. Posterior distributions suggest that the split between eastern and western foxsnakes occurred approximately 9817 years in the past (90% confidence interval of 5317–14 587 years ago). Again this timing strongly suggests that the disjunctions did not result from European settlement but predated the proposed timing of the infilling of deciduous forest. The wide confidence intervals and low R^2 suggest, however, that we may not have significant power to estimate this splitting time with our microsatellite markers alone and when estimating the splitting time from a simple two-population model, the 90% highest posterior density estimate (1732–13 185 years) would have included the infilling of deciduous forest. It is also possible that generation time for more southerly populations would be shorter, which would decrease these timing estimates.

Anthropogenic habitat alteration and conservation implications

Although the large population declines and regional population splits probably occurred before major European colonization, we found some evidence that agricultural, residential and urban development have further impacted populations across the distribution, but at finer geographical scales. Indeed, Row *et al.* (2010) found that disjunctions between diagnosed genetic clusters in southwestern Ontario correlated well with agricultural fields and road barriers. The timing of the population split in this region (20–250 HPD generations; 150–1875 years) is consistent with the notion that anthropogenically driven habitat fragmentation isolated previously larger and more connected populations of foxsnakes in this region. Results from our ABC analysis also implied that the current population sizes of foxsnakes are much smaller than those in the past, which is especially true for eastern foxsnakes. Although it appeared that the largest decline predated extensive European settlement, it is unlikely that anthropogenic habitat loss and fragmentation is not continuing to affect populations, as evidenced by the southwestern Ontario analysis. There is recent evidence of a widespread decline in snakes (Reading *et al.*, 2010), and small increases in mortality can have large impacts on populations of late maturity species, such as large snakes in temperate climates (Row *et al.*, 2007). Ultimately, combining these population size estimates with population viability analyses would be beneficial for determining the viability of these remaining populations.

Conclusions

Our study provides a firm foundation for future work both on foxsnakes and on co-distributed species. Schmidt (1938) used the eastern range extension of 11 primarily prairie distributed species of herpetofauna as evidence for the post-glacial steppe. Other studies have since identified similar ranges of other species of herpetofauna, as well as species of mammals, plants and insects (Thomas, 1951; Smith, 1957; Lloyd, 1967). Many of these species are also associated with wetland habitats (e.g. turtles and frogs) and likely also benefited from the lake formation and drainage basins from the

melting ice caps (Mockford *et al*., 2007). Similar tests of the post-glacial expansion of these other species would reveal whether they show evidence for population declines that are similar in timing and extent to those of foxsnakes. An examination of massasauga rattlesnakes (*Sistrurus c. catenatus*) would be particularly useful, as their range in Ontario is similar to eastern foxsnakes with disjunct populations in southwestern Ontario and the Georgian Bay area.

The Approximate Bayesian computing approach (Beaumont *et al*., 2002, 2010) that we deployed provided a robust hypothesis-testing framework for comparing alternate historical demographic models. In our hierarchical analysis, the single-population models and regional population models showed consistent results in terms of splitting times and population sizes. Also, when combining eastern and western foxsnakes into populations separately, the splitting time estimate was relatively consistent with the complex models, but seemed to produce an unrealistic estimate for the effective population size of eastern foxsnakes (e.g. N_e of 2000 for the entire range). Comparing estimates between models increased our confidence in the results, but also suggested that ABC analysis may be robust in some situations where there are gaps in sampling and/or when complex situations are simplified. Simulation studies evaluating the impact of incomplete sampling and simplifying complex demographic scenarios would be particularly fruitful.

Acknowledgments

For their hard work and assistance in the field, we would first like to thank H. Row, C. Hudson, K. Geale, R. Lougheed, K. Vincent, K. Donmoyer, N. Morrill, C. Monk and A. Xuereb. For generously collecting and/or providing blood and tissue samples, we thank K. Stanford, K. Bekker and B. Putman, G. Nelson, K. Kucher, B. Groves, D. Jacobs, R. Gould, D. Hector, V. McKay, Chicago Field Museum, the staff at the Ojibway Nature Center and at Point Pelee National Park. For assistance with python scripting used in the ABC analysis, we also thank B. Chau. R. King, R. Montgomerie and V. Friesen provided constructive criticism of an earlier incarnation of this manuscript. For financial support and/or providing collection permits, we gratefully acknowledge the support of World Wildlife Fund (through the Endangered Species Recovery Fund), Environment Canada, Ontario Ministry of Natural Resources, Parks Canada, the Essex County Stewardship Network, Shiawassee National Wildlife Refuge, U.S. Fish and Wildlife Service, NSERC (NSERC Discovery Grant to SCL and postgraduate scholarship to JRR) and Queen's University through the Summer Work Experience Program (SWEP). We especially thank K. Prior, A. Promaine J. Rouse and T. Dobbie for their assistance and support throughout this project.

References

Aldenhoven, J., Miller, M., Corneli, P. & Shapiro, M. 2010. Phylogeography of ninespine sticklebacks (*Pungitius pungitius*) in North America: glacial refugia and the origins of adaptive traits. *Mol. Ecol.* **19**: 4061–4076.

Austin, J.D., Lougheed, S.C., Neidrauer, L., Chek, A.A. & Boag, P.T. 2002. Cryptic lineages in a small frog: the post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Mol. Phylogenet. Evol.* **25**: 316–329.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E. *et al.* 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**: 489–522.

Beaumont, M. 2010. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* **41**: 379–406.

Beaumont, M.A., Zhang, W. & Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.

Beaumont, M.A., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L. *et al.* 2010. In defence of model-based inference in phylogeography. *Mol. Ecol.* **19**: 436–446.

Bertorelle, G., Benazzo, A. & Mona, S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* **19**: 2609–2625.

Blouin-Demers, G. & Gibbs, H.L. 2003. Isolation and characterization of microsatellite loci in the black rat snake (*Elaphe obsoleta*). *Mol. Ecol. Notes* **3**: 98–99.

Burbrink, F.T., Lawson, R. & Slowinski, J.B. 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution* **54**: 2107–2118.

Chikhi, L., Sousa, V., Luisi, P., Goossens, B. & Beaumont, M. 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186**: 983–995.

Churcher, C. & Karrow, P. 2008. The Hamilton Bar Fauna: evidence for a Hypsithermal age. *Can. J. Earth Sci.* **45**: 1487–1500.

Collins, J.T. 1991. Viewpoint: a new taxonomic arrangement for some North American amphibians and reptiles. *Herpetol. Rev.* **22**: 42–43.

Collins, J. & Taggart, T. 2008. An alternative classification of the New World rat snakes (genus *Pantherophis* [Reptilia: Squamata: Colubridae]). *Herpetology* **26**: 16–19.

Conant, R. 1940. A new subspecies of the Fox Snake, *Elaphe vulpina* Baird and Girard. *Herpetologica* **2**: 1–14.

Conant, R. & Collins, J.T. 1991. *A Field Guide to Amphibians and Reptiles of Eastern and Central North America*. Houghton Mifflin Co., Boston, Massachusetts.

Cornuet, J.M., Ravigné, V. & Estoup, A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *Bioinformatics* **11**: 401.

Crawford, N.G. 2010. SMOGD: software for the measurement of genetic diversity. *Mol. Ecol. Resour.* **10**: 556–557.

Csillery, K., Blum, M.G.B., Gaggiotti, O.E. & Francois, O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**: 410–418.

Dieringer, D. & Schlotterer, C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**: 167–169.

Dileo, M.F., Row, J.R. & Lougheed, S.C. 2010. Discordant patterns of population structure for two co-distributed snake species across a fragmented Ontario landscape. *Divers. Distrib.* **16**: 571–581.

Dyer, R., Nason, J. & Garrick, R. 2010. Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Mol. Ecol.* **19**: 3746–3759.

Eckert, C.G., Samis, K.E. & Lougheed, S.C. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Mol. Ecol.* **17**: 1170–1188.

Ehlers, J. & Gibbard, P.L. 2004. *Quaternary Glaciations: Extent and Chronology*. Elsevier, Amsterdam, The Netherlands.

Estoup, A. & Clegg, S. 2003. Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Mol. Ecol.* **12**: 657–674.

Excoffier, L. & Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**: 564–567.

Excoffier, L., Estoup, A. & Cornuet, J.M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727–1738.

Fagundes, N.J.R., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L. *et al.* 2007. Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. USA* **104**: 17614–17619.

Garza, J.C. & Williamson, E.G. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**: 305–318.

Gibbs, H.L., Corey, S., Blouin-Demers, G., Prior, K. & Weatherhead, P.J. 2006. Hybridization between mtDNA-defined phylogeographic lineages of black ratsnakes (*Pantherophis* sp.). *Mol. Ecol.* **15**: 3755–3767.

Gould, S.J. & Johnston, R.F. 1972. Geographic variation. *Annu. Rev. Ecol. Syst.* **3**: 457–498.

Hollander, M. & Wolfe, D.A. 1999. *Nonparametric Statistical Methods*, 2nd edn. John Wiley & Sons, New York, NY.

Howes, B.J. & Lougheed, S.C. 2008. Genetic diversity across the range of a temperate lizard. *J. Biogeogr.* **35**: 1269–1278.

Howes, B.J., Lindsay, B. & Lougheed, S.C. 2006. Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Mol. Phylogenet. Evol.* **40**: 183–194.

Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577–586.

Jakobsson, M. & Rosenberg, N. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Johansson, M., Primmer, C.R. & Merilae, J. 2006. History vs. current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). *Mol. Ecol.* **15**: 975–983.

Jombart, T. 2008. *adegenet:* a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Jost, L. 2008. GST and its relatives do not measure differentiation. *Mol. Ecol.* **17**: 4015–4026.

Joyce, P. & Marjoram, P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* **7**: 26.

King, J.E. 1981. Late Quaternary vegetational history of Illinois. *Ecol. Monogr.* **51**: 43–62.

Knowles, L.L. & Alvarado-Serrano, D. 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Mol. Ecol.* **19**: 3727–3745.

Knowles, L.L. & Maddison, W.P. 2002. Statistical phylogeography. *Mol. Ecol.* **11**: 2623–2635.

Laval, G. & Excoffier, L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**: 2485–2487.

Leuenberger, C. & Wegmann, D. 2010. Bayesian computation and model selection without likelihoods. *Genetics* **184**: 243–252.

Lloyd, J. 1967. The Prairie Peninsula and secondary intergradation in *Photinus* fireflies (Coleoptera: Lampyridae). *Coleopt. Bull.* **21**: 33–39.

Manel, S., Gaggiotti, O.E. & Waples, R.S. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol.* **20**: 136–142.

Mockford, S.W., Herman, T.B., Snyder, M. & Wright, J.M. 2007. Conservation genetics of Blanding's turtle and its application in the identification of evolutionarily significant units. *Conserv. Genet.* **8**: 209–219.

Morse, M. 1902. The range of the fox snake. *Science* **15**: 1035.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Neuenschwander, S., Largiadèr, C.R., Ray, N., Currat, M., Vonlanthen, P. & Excoffier, L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol. Ecol.* **17**: 757–772.

Nielsen, R. & Beaumont, M.A. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* **18**: 1034–1047.

Panchal, M. & Beaumont, M.A. 2007. The automation and evaluation of nested clade phylogeographic analysis. *Evolution* **61**: 1466–1480.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**: 1791–1798.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Qu, Y., Lei, F., Zhang, R. & Lu, X. 2010. Comparative phylogeography of five avian species: implications for Pleistocene evolutionary history in the Qinghai-Tibetan plateau. *Mol. Ecol.* **19**: 338–351.

R Development Core Team 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ray, N., Wegmann, D., Fagundes, N.J.R., Wang, S., Ruiz-Linares, A. & Excoffier, L. 2010. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol. Biol. Evol.* **27**: 337–345.

Reading, C.J., Luiselli, L.M., Akani, G.C., Bonnet, X., Amori, G., Ballouard, J.M. *et al.* 2010. Are snake populations in widespread decline? *Biol. Lett.* **6**: 777–780.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**: 137–138.
Row, J.R., Blouin-Demers, G. & Weatherhead, P.J. 2007. Demographic effects of road mortality in black ratsnakes (*Elaphe obsoleta*). *Biol. Conserv.* **137**: 117–124.
Row, J.R., Sun, Z., Cliffe, C. & Lougheed, S.C. 2008. Isolation and characterization of microsatellite loci for eastern foxsnakes (*Elaphe gloydi*). *Mol. Ecol. Resour.* **8**: 965–967.
Row, J.R., Blouin-Demers, G. & Lougheed, S.C. 2010. Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. *Mol. Ecol.* **19**: 5157–5171.
Schmidt, K. 1938. Herpetological evidence for the postglacial eastward extension of the steppe in North America. *Ecology* **19**: 396–407.
Schoville, S. & Roderick, G. 2009. Alpine biogeography of Parnassian butterflies during Quaternary climate cycles in North America. *Mol. Ecol.* **18**: 3471–3485.
Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* **236**: 787–792.
Smith, P. 1957. An analysis of post-Wisconsin biogeography of the Prairie Peninsula region based on distributional phenomena among terrestrial vertebrate populations. *Ecology* **38**: 205–218.
Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**: 381–397.
Thomas, E. 1951. Distribution of Ohio animals. *Ohio J. Sci.* **51**: 153–167.
Webb, T. 1981. The past 11,000 years of vegetational change in eastern North America. *Bioscience* **31**: 501–506.
Wegmann, D. & Excoffier, L. 2010. Bayesian inference of the demographic history of chimpanzees. *Mol. Biol. Evol.* **27**: 1425–1435.
Wegmann, D., Leuenberger, C. & Excoffier, L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**: 1207–1218.
Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**: 116.
Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
Wright, S. 1978. *Evolution and the Genetics of Populations, Volume IV: Variability within and among Natural Populations*. University of Chicago Press, Chicago, IL.
Zar, J.H. 1996. *Biostatistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersy, NY.
Zellmer, A.J. & Knowles, L.L. 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Mol. Ecol.* **18**: 3593–3602.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Mitochrondrial structure and diversity.

Appendix S2 Posterior predictive distribution for selected models.

Appendix S3 Effect of Gene Flow on parameter estimates.

Figure S1 Biplots of individual genotypes for (a) PCA axis 1 (*x*-axis) versus PCA axis 2 (*y*-axis) and, (b) PCA axis 1 (*x*-axis) versus PCA axis 3 (*y*-axis).

Table S1 Sample sizes and specific locations for sampled populations across the range of eastern and western foxsnakes.

Table S2 Prior distribution ranges used for parameters in Approximate Bayesian computation models (Illustrations of models in Fig. 2a.) designed to estimate the demographic history of foxsnake populations.

Table S3 Prior distribution ranges used for parameters in Approximate Bayesian computation models (Illustrations of models in Fig. 2.) designed to estimate the demographic history of foxsnake populations in southwestern Ontario.

Table S4 Prior distribution ranges used for parameters in Approximate Bayesian computation models (Illustrations of models in Fig. 3) designed to estimate the demographic history of foxsnake populations across their range.

Table S5 Pairwise F_{ST} values (bottom) and Jost's D differentiation values (top) between genetic clusters (see Fig. 1 for population distributions). All pairwise F_{ST} values were highly significant ($P < 0.001$).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 4 April 2011; revised 9 June 2011; accepted 28 June 2011